DAVIDSON'S TEST

Donald Davidson's Critique of the Turing Test as an Expression of his Theory of Intellectual and Linguistic Competence

TOBIAS LOHSE

OCTOBER 22, 2015

Abstract

In his essay 'Turing's Test' ([1990b] 2004), Donal Davidson discusses Alan Turing's famous proposal how to define the goal of Artificial General Intelligence and proposes a modified version that fixes the weaknesses of the original proposal. Davidson's Test illustrates his theory about the essence of linguistic and intellectual competence very well.

Davidson's idea of triangular Epistemology

Keywords

Donald Davidson, Philosophy of Mind, Philosophy of Language, Epistemology, Interpretation, Turing Test, Artificial Intelligence

Contents

1 Introduction **3**

1.1 Introduction to Donald Davidson 3

1.2 Overview of the Paper 3

2 The Turing Test and Its Relevance **3**

2.1 Interpretations and Versions of the Turing Test 3

2.2 Relevance of the Test and the Turing Principle 3

2.3 Turing on Machine Learning 4

3 Davidson's Critique of the Turing Test **5**

3.1 Relevance of Turing's Test for Davidson 5

3.2 Davidson's Critique of the Turing's Test 6

3.3 Davidson's Proposal for a Modified Test 6

4 Conditions for Linguistic Competence **6**

4.1 The Distinction between Linguistic Competence and Performance 6

4.2 Recursiveness and Empirism 6

4.3 The Social Aspect of Language 6

4.4 Anti-Conventionalism 6

4.5 The Triangular Nature of Knowledge 6

5 Davidson's Theory of Intellectual and Linguistic Competence **7**

5.1 Sources for Davidson's Theory of Intellectual and Linguistic Competence 7

5.2 Outline of the Unified Theory 7

5.3 Relation to Davidson's Test 7

5.4 Possible Objections against Davidson's Test 7

6 Conclusion **7**

6.1 Prospects of Davidson's Pragmatic Theory 7

6.2 Implications for Artificial Intelligence and Computer Linguistics 8

References **8**

1. Introduction

1.1. Introduction to Donald Davidson

1.2. Overview of the Paper

2. The Turing Test and Its Relevance

Alan Turing (1912-1954) first proposed his famous test in 'Computing Machinery and Intelligence' (1950). The thesis of his paper is that the question "Can machines think?" — which he deemed too vague to deserve discussion — can be replaced with the question wether a computer can pass a specified test (that is now known as the 'Turing Test'). This test is described as a game: the imitation game.

The imitation game consists of an *interrogator* who can communicate via a real-time text chat interface with two players. One player is a *computer*, the other a *human*. The interrogator's task is to identify the human after a given time. The computer's task is to pretend to be a human[1] and trick the interrogator into wrongly identifying it as the human. The human's task is to help the interrogator to make the correct identification. If many interrogators consistently[2] can't tell the computer apart from the human the computer wins the game and passes the test. (Turing 1950, p. 433–434)

In the following I discuss different interpretations of the Turing Test and establish which interpretation and version of the test I take as the basis for our discussion. Furthermore I argue for the relevance of this interpretation of the test and work out 4 main claims from Turing's discussion of his test. In doing so I emphasize some of Turing's visionary ideas on how a computer must be programmed to pass the test, that have implications for Davidson's discussion of the test.

2.1. Interpretations and Versions of the Turing Test

The interpretations are not as straight forward, as it might seem. There are three main interpretations for what it means to pass the Turing Test:

1. The Turing Test is taken as an operational *definition* — a *sufficient* and *necessary* condition — of *intelligence*[3]: Something is intelligent if and only if it passes the Turing Test. (Such an interpretation is for example found in Millar 1973.)
2. The Turing Test is taken as a *sufficient* condition of intelligence: Something is intelligent if it passes the Turing Test. But it is not necessary for something to pass the Turing test in order to be intelligent. (Donald Davidson [1990b] 2004 tends towards this interpretation.)
3. The Turing Test is taken "as a potential source of good inductive evidence for" intelligence: If Something passed the Turing Test one would be *justified* for inductively inferring that it is intelligent. (This interpretation goes back to Moor 1976, p. 249, 251.)

The first interpretation and any reading of intelligence in a broader sense are clearly not in line with Turing's ideas. Turing was not interested in a definition of intelligence but in setting a clear goal for further research. In fact I would argue that the third interpretation from Moor does Turing most justice. This becomes fairly clear from the following section of a radio interview ([1952] 1999), which represents Turing's simplest expression of the test:

> "I *don't want to give a definition of thinking* [...] I don't really see that we need to agree on a definition at all. The important thing is to try to *draw a line between the properties of a brain, or of a man, that we want to discuss, and those that we don't.* [...] I would like to suggest a particular kind of test that one might apply to a machine. You might call it a test to see whether the machine thinks, but it would be better to avoid begging the question, and *say that the machines that pass are (let's say) 'Grade A' machines*. The idea of the test is that the machine has to try and pretend to be a man, by answering questions put to it, and it will only pass if the pretence is reasonably convincing."*(Turing [1952] 1999, p. 466, emphasis added)

Turing clearly describes the imitation game as a test that defines a special *class of computers*, but not as one that defines thinking/intelligence. It is also implicit that he sees the *ability to communicate* as quintessential for human-like intellectual abilities.

The quotation also gives a simplified version of the test, in so far as it removes the second human player. While this might be limiting for a quantitative analysis, it ensures the focus on the main question and removes any undue emphasis on strategy. Davidson proposes the same simplification in his discussion ([1990b] 2004). He goes even further in proposing that "the interrogator [...should simply] be asked to decide whether or not the object is thinking" ([1990b] 2004, p. 81). If we assume that any bias against the object's ability to think can be removed (which makes sense for this philosophical discussion), this seems to be the clearest expression of Turing's idea. Therefore, I will follow Davidson and refer to this simplified version in the following.

2.2. Relevance of the Test and the Turing Principle

Where does it leave the relevance of the Turing Test, if we follow Moor and interpret it merely as a framework to gather evidence for human-like intelligence? For Computer Scientists it might be *prudent* to follow Turing and ignore the philosophical question of a definition of intelligence and instead take the Turing Test as a definition of a certain class of computers. In this way the Turing Test can give a clear

[1] I think it is pretty clear that Davidson ([1990b] 2004, p. 78) misinterprets Turing when he suggests that Turing wants the computer to play the literal imitation game and pretend to be a *woman* and that the interrogator's task would be to decide on the gender. This becomes pretty clear when Turing says: "If the man were to try and pretend to be the machine [...]" (1950, p. 435) and from all his examples that are focused on how the machine can imitate being a *human* not a woman. (See also Copeland 2000, p. 526.)

[2] Turing's idea is that a baseline is established by the traditional imitation game, in which a man tries to imitate a woman and the interrogator has to decide on the gender. While this might not be the best way to establish a baseline, it means that Davidson's critique that "Turing does not say what he would make of a computer that was consistently chosen over the [...human] to be the [...human]" ([1990b] 2004, p. 78), is ill-conceived, as the computer will always unambiguously fall over or under the baseline within a margin of error.

[3] The term 'intelligence' is adopted here purely in reference to Artificial Intelligence. Turing mostly uses the term 'thinking' and Davidson (1995) mostly uses the term 'rationality' to refer to the concept that this thesis is concerned with. I will introduce and explicate the term intellectual competence for this concept in section 4.1.

[4] See Muehlhauser (2013) for a brief overview of operational definitions of AGI, including the Coffe-Brewing and College Test. From a philosophical perspective all those seem rather random and dubious — certainly much further away from clearly capturing necessary conditions of human-like intellectual competences.

goal to the field of 'Artificial General Intelligence' (AGI), for which the unclarity about a clear definition of its topic is a core challenge. Unlike other propositions for an operational definition[4] of AGI, the Turing Tests provides a clear and well-justified empirical goal. In his analysis Moor argues why the Turing Test is well apt for that:

> "[T]here are two strong arguments why the Turing test is a good format for gathering inductive evidence. First, the Turing test permits direct or indirect testing of virtually all of the activities one would count as evidence for thinking. Secondly, the Turing test encourages severe testing. [...T]he computer would be tested in detail over a wide range of subjects [...and] the interrogator's goal is to find a refuting instance which gives the computer away." (Moor 1976, p. 251–252)

Critics of the Turing Test as a goal for AGI mostly fall into two camps. The first point Moor provides is aimed against the first type of critic who suggests the test sets the wrong goal. As Moor argues, communication is a very clear framework to investigate all kinds of thinking. A further practical reason for the relevance of the Turing Test — that I would add — is that natural language communication provides a 'gold standard' for completely natural and seamless computer user interfaces. Many of those critics mistake the question as philosophical while it is best treated as definitional. Computer Scientists ought not to be concerned with defining intelligence in general but with a good definition for "Grade A" computers — to use Turing's terminology. In this regard, the interpretation of the Turing Test as a framework to gather empirical evidence can be set as a clear definition of the goal of AGI.

The other type of critic questions the adequacy of the test to determine whether the goal is reached. Moor's second point is aimed against that by pointing out how well the test encourages thorough testing. The critics sometimes mistake Turing's predictions as a specification that the test ought to take only 5 minutes and limited implementations of the Turing Test that favor engineering tricks like the Loebner Prize[5] discredit Turing's intentions. I agree with Copeland's interpretation of Turing: the goal set by the test is a computer that "plays the imitation game successfully come what may, with no field of human endeavour barred, and for any length of time commensurate with the human lifespan." (Copeland 2000, p. 530)

TURING'S THREE MAIN CLAIMS

Following (Copeland 2000, p. 530), we can establish the following two main claims of Turing from our previous discussion:

1. Turing sees communication as suitable to expose the relevant intellectual abilities that determine wether a computer can perform tasks on paar with a human being.
2. Turing thinks that his test *specifically* can determine wether a computer possesses such communicative abilities.

But there is also a third claim which is known as the Turing Principle[6]:

3. Turing believes that universal computers can simulate any physical process, including the brain.

This might be most clearly expressed in Turing's lecture on 'Can Digital Computers Think?': "If it is accepted that real brains [...] are a sort of machine it will follow that our digital computer suitably programmed will behave like a brain" (Turing [1951] 1999, p. 463). This is a much more controversial claim and — as Turing was well aware — he had few arguments and even less evidence for it.

It is important to highlight that this claim is not central in Turing's writing and that it is separate and fully *independent* from the others claims. Nevertheless, it is the main point many philosophers have attacked. (The most famous example is probably Searle 1980.) This is often tied to the misinterpretation of the Turing Test as a definition or sufficient condition for intelligence in a philosophical sense. I am not particularly interested in the discussion of this third claim here.

2.3. Turing on Machine Learning

Unlike the obsession in popular culture with human-like 'Artificial Intelligence' (AI) might suggest, the interest of Computer Scientists in the last years has been more focused on so-called 'weak' AI. This refers to applying 'Machine Learning' to apply domain-specific intelligence to special problems (most notably: image recognition, domain-specific language processing, and robotics). The field that is concerned which this is referred to as 'Artificial Intelligence' these days. The field that is concerned with human-like strong AI is the field of Artificial General Intelligence (AGI) which was referenced in the previous section. This shift of focus has mainly been due to a lack of success in AGI.

Turing believed that by the turn of the century there would be computers that "play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning" (1950, p. 442). Recent winners of the Loebner Prize (AISB, n.d.), which is awarded to the most human-like chatbot judged by a Turing-inspired test, show that this has not quite come to pass. On the other hand, Turing himself said that it would take "*at least* 100 years" ([1952] 1999, p. 434) until a general Turing Test could be passed — and the goal of AGI reached — which is still well within the realm of the possible. Turing's wrong prediction of the development can be mainly accredited to his underestimation of the required processing power for Machine Learning (1950, p. 455). Recent accomplishments in AI with deep neural networks, for example, have only become feasible because computers are able to run networks with millions of neurons in real time.[7]

[5] Shieber (1994) criticizes the Loebner prize for its inappropriateness to award advances in natural-language-processing techniques instead of engineering tricks oriented to the exigencies of the restricted task like parrying and insertion of random typing errors. But the setup of the scoring system alone shows how pointless it is to even judge current systems by a direct Turing Test.

[6] This is also known as the 'Church–Turing–Deutsch Principle' and represents an extension of the well-known Church-Turing Thesis to artificial intelligence.

[7] Steven Wolfram writes: "Computers (and especially linear algebra in GPUs) got fast enough that [...] it became practical to train neural networks with millions of neurons, on millions of examples. [...T]his suddenly brought large-scale practical applications within reach. [...] I don't think it's a coincidence that this happened right when the number of artificial neurons being used came within striking distance of the number of neurons in relevant parts of our brains. [...I]f we're trying to achieve 'human-like' image identification [...] then this defines a certain scale of problem, which, it appears, can be solved with a 'human-scale' neural network." (2015)

TURING'S FOURTH MAIN CLAIM

However, Turing's idea about *how* to build a computer that could pass his test are more interesting than his timeline predictions. His claim that "the problem [of building a computer that passes the Turing Test] is mainly one of programming" (1950, p. 455) still rings true today.

We have already learned of his idea that computers should be able to simulate the brain ([1951] 1999). Indeed this reverse engineering approach to AI is the basic idea behind the neural networks which power the most successful image entity-recognition algorithms today and was already investigated by Turing ([1948] 1992; see also Copeland and Proudfoot 1999). It seems plausible that we might achieve AGI through brain simulation before we deeply understand how the brain works.

Even more interesting is that Turing also predicted the Machine Learning approach to build intelligent algorithms and saw its non-deterministic nature as a characteristic feature:

> "Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain. [...] We have thus divided our problem into two parts. The child-programme and the education process. These two remain very closely connected. We cannot expect to find a good child-machine at the first attempt. One must experiment with teaching one such machine and see how well it learns." (Turing 1950, p. 456)

> "An important feature of a learning machine is that its teacher will often be very largely ignorant of quite what is going on inside, although he may still be able to some extent to predict his pupil's behaviour. [...] This is in clear contrast with normal procedure when using a machine to do computations : one's object is then to have a clear mental picture of the state of the machine at each moment in the computation. [...] Processes that are learnt do not produce a hundred per cent certainty of result; if they did they could not be unlearnt." (Turing 1950, p. 458–459)

This description is very much how modern statistical Machine Learning algorithms work. They consist of an algorithm that describes a mathematical model which is trained with human-annotate data (for example, a grammar analysis of sentences) and are then able to perform their task on similar data. (See Schubert 2015 for an overview about approaches to Natural Language Processing; and Jurafsky and Martin 2015 for a more technical introduction.) As Turing mentions, this is a paradigm shift from classical algorithms which have results that are clearly defined by the programmer and predictable independent of any training data. We might add a fourth core claim:

4. Turing believes that the approach to device intelligent computers must be based on learning algorithms that are trained and do not behave predictable in a classic sense.

Note that devising intelligent algorithms is a quite different task and involves techniques where an interpretation of the state of the program at each step becomes difficult or even impossible — especially in the case of deep neural networks. And where the outcome is not just dependent on the set of rules specified in the programming but also on the 'experience' gathered in the program during its learning process. We will see later why this matters for a Davidsonian perspective.

SUMMARY

We have seen that a simplified interpretation of the Turing Test as proposed by Donald Davidson ([1990b] 2004) reveals the core of Turing's idea. However, we have also seen that an interpretation of the Turing Test as an operational definition or sufficient condition for intelligence is not in line with Turing's writing. Instead, I have proposed to follow Moor (1976) and interpret the test as a framework to collect empirical evidence to show that a computer can perform human-like tasks. We have seen that this can provide a good and clear definition of a special class of computers that are the goal of AGI research.

Furthermore, we have established that Turing claims that (1) communication abilities are representative of intellectual abilities in general and (2) that his test is adequate to evaluate those abilities. I have pointed out that his claim (3) that computers can simulate the brain (the Turing Principle), which is the main point most philosophers critique, is completely independent of the other claims and is the least essential one for Turing.

Lastly, we have seen that Turing had pioneering ideas about how to devise algorithms which could pass his test. Such algorithms (4) need to be able to learn when trained with data and, different from classical algorithms, their outcome is not predictable and the states of their operation are not easily interpretable.

3. Davidson's Critique of the Turing Test

3.1. Relevance of Turing's Test for Davidson

Why the test is interesting to Davidson - Test is concerned with the nature of thought - Pragmatic approach to the question of thought - Empiric criterion for thought - Not confined to maschines

> *"The whole thinking process is still rather mysterious to us, but I believe that the attempt to make a thinking machine will help us greatly in finding out how we think ourselves."* (Turing [1951] 1999, p. 465)

"the test is designed to throw light on the nature of thought." (Donald Davidson [1990b] 2004, p. 78)

> *"I believe that another human being thinks because his ability to think is part of a theory I have to explain his actions. [...] the evidence for the theory comes from the outward behavior of the person. [...] there is no reason why knowledge of computer thinking can not arise in the same way. I can use the computer's behavior as evidence in assessing my theory about its information processing. In neither the human case nor the computer case must I consider the thinking to be on a close analogy with my own, for the evidence might dictate that the human or computer discriminates and evaluates quite differently than I do."* (Moor 1976, p. 251)

Where does Davidson agree - There should be an empiric test for intellect - Linguistic comptetence is essential for intellect - General/Strong AI might well be possible - Davidson sees no argument why AI shouldn't be possible - AI through Brain simulation should be possible - See also 'Representation and Interpretation' (Donald Davidson [1990a] 2004) - No need for introspection into working of mind (against Searle's chinese room)

Turing's proposed test shows that he agrees with Davidson on the fact that linguistic competence is essential to intellect. And I think we also have to interpret Davidson and Turing as agreeing that there is a scientific approach that can describe essential parts of our linguistic competence. In the case of Turing this is evident in his believe that there is a program that allows the computer to win the imitation game. For Davidson it is evident in his proposal of the empiric Unified Theory that can capture the essence of linguistic competence and rationality. We will investigate Davidson's ideas further in [Unified Theory of Understanding].

3.2. Davidson's Critique of the Turing's Test

The core argument: knowledge comes from interaction in shared world - Rejection of sharp distinction between mental and physical (especially sensory) abilities (E4.5) - distinction between semantic and syntactic abilities (+E5.8) - Relation to Agreement between Descartes and Turing - Both come from the idea that knowledge is simply in the head - Even Turing's Helen Keller example lacks as she was interacting with the world through touch. - Connected to rejection of Cartesion Epistemology - relation to French (1990)

Does Davidson's Critique Do Justice to Turing? - Davidson has a good point - but he doesn't do justice to Turing's thoughts on Machine Learning that would deserve a discussion in his paper - His emphasis on a history of causal interaction with the world is partly satisfied by machine learning

APPROPRIATENESS OF THE CRITIQUE

3.3. Davidson's Proposal for a Modified Test

Observe a History of interaction with the world - The ability to determine non predefined meaning is essential for intelligence

Summary

4. Conditions for Linguistic Competence

4.1. The Distinction between Linguistic Competence and Performance

Distinction between sufficient and necessary conditions of linguistic competence - relation to term linguistic competence and linguistic performence - analogy to physics not being able to describe actual natural systems but only systems under ideal conditions - See (E4.8, E5.8)

I use the term *linguistic competence*[8] here to refer to Davidson's idea of the essential aspect of understanding language vs. the practical understanding of language.

Relation to Davidson's Philosophy - truth conditions as a formal language - Unified Theory, Radical Interpretation - Sucessfull Communication & Anti-Conventionalism & Holims of knowledge

4.2. Recursiveness and Empirism

The Importance of Empirical Theories - No a priori knowledge about language(-learning), only empirical knowledge (E2.1)

Learnable Languages - Finite set of linguistic primitives gives rise to infinite linguistic expressions

'Theories of Meaning and Learnable Languages' (Davidson 1965)

4.3. The Social Aspect of Language

The Primacy of Ideolect - relevance of interpretation rather than meaning (+E2.2) - see similarities to Gadamer's Hermeneutics of Wirkungsgeschichte (Bertram 2012) - Relevance of Intention in communication (+E5.10) - Arguments for primacy of ideoloect over language - Malapropisms and the like (E5.7)

Why is language necessarily Social - Wittgenstein's norm of actually getting it right in language - also see language game analogy in anti conventionalism (2.18) - and importance of disctinction between true and false beliefs in triangulation (E3.14) - Truth only determined trough communication - see also Plato, socrates (E5.16)

4.4. Anti-Conventionalism

arguments against conventionalism - impossibility of public sign for sincerity - language game analogy (+E5.8)

Remarks about how understanding does work - At the end of (1984), Davidson says that understanding can't be formally described - Relation to distinction linguistic between competence and prerformence

4.5. The Triangular Nature of Knowledge

The basic problem of Epistemology - Distinctiveness of three types of Knowledge (subjective, intersubjective, objective) (E3.14) - Three Questions of Epistemology (E3.14) - how to know other minds - how to know world - how to know own mind without evidence

Rejection of other Epistemologies - Rejection of Positivism (reductionism to world knowledge) (E3.14) - Rejection of empiricsit Epistemology (E3.14, E2.1) - No primacy of names and predicates in the foreground of senses - doesn't he contradict that in E2.9 and E4.8? - Rejection of Cartesion Etymology (primacy of knowledge of one's own mind) (E3.14, E3.3, E3.13) - Rejection of skepticism (E3.14, SEP Davidson, E4.1)

[8] The term linguistic competence was introduced by Chomsky who makes a "fundamental distinction between competence (the speaker-hearer's knowledge of his language) and performance (the actual use of language in concrete situation)" (Chomsky 2014, p. 4). While Davidson criticizes the idea of linguistic competence as knowledge of a language, he adopts the term for the description of his theory that describes the interpreters essential competence in 'A Nice Derangement of Epitaphs' (1986).

The Process of Gaining Knowledge - Interdependency of three varieties of Knowledge - Importance of disctinction between true and false beliefs (E5.8, E3.14, E4.1, SEP Davidson) - Principle of Charity (Coherence and Correspondence) (E3.14, E2.9, E3.14) - Reaction to stimuli (E3.14) - proximal vs. distal stimuli (E5.4) - Iscrutability of Reference / Indeterminacy of meaning (Quine Word and Object, E5.3, E2.16, E3.14)

The Underlying Ontology and Metaphysics - Ontology of only objects and events (E3.14, E5.4) - Directness of knowledge of world - no intermediary entities (5.3, E5.4, E3.10) - no mentales, language only through communication, meaning given causally by objects events (E5.9) - Truth as elementary non-reducible concept (E5.2, E2.I)

Anomalous Monism - parallels to Spinoza's causal theory of affects (E5.20) - see arguments for irreducibility of mental in Tirangulation (E3.14) - Reasons for Irreducibility/anaomalism of Mental to Physical (E3.14, E4.8) - Irreducibility of knowledge of belief to knowledge of world - causal nature of the mental (+E1.11, E4.8) - Missing independent mode of communication about mental theories (+E4.8) - Turing's discussing of rule based behavior

Turing discusses an interesting objection to his test that is related to Anomalous Monism (1950, p 452f). If our behavior is not ruled by strict laws but a computer is. How could it ever simulate our behavior. Turing argues that this is simply an assumptions (and even if who says that the mental states of a computer could for example not supervene on the sates of its neural network). Turing uses cryptography as an example for a process where a computer can produce output which can not be tracked back to its originating rules (but this of course is a practical limitation not a theoretical one, since all cryptographic functions can be reversed with infinite processing power.)

This raises and important question wether the anomalism of the mental is connected to our linguistic competences (or maybe just to our effectiveness?) and how computers/the unified theory could account for it. ***(need to reread in three varieties of knowledge)***

Summary

5. Davidson's Theory of Intellectual and Linguistic Competence

5.1. Sources for Davidson's Theory of Intellectual and Linguistic Competence

What is this theory of interpretation - Anomalous Monism - Davidson's Program - Radical Interpretation - Triangulation - Unified theory of Thought, Action, and Meaning - Prior and Passing Theories - The social Aspect of language

So what is it that I am referring to when I speak about Davidson's theory of interpretation? While this seems like a straight forward question it is a little more complicated because of the distributed nature of Davidson's work.

Davidson's theory can be traced back to his rejection of meaning and idea that interpretation can be described as a process of 'translating' utterances into truth conditions based on assumptions about the believes of the speaker using the principle of Charity, which he develops in 'Truth and Meaning'[9] (Davidson 1967) and 'Radical Interpretation' (Donald Davidson 1973b).

5.2. Outline of the Unified Theory

Decision Theory

Representation of Meaning/Linguistic Content - Meaning as Language (E3.14, E2.2) (not things in world, not entities sui generis)

How does the interpretation Process work

5.3. Relation to Davidson's Test

Interpreting Davidson's Test as a Test for a Unified Theory 'machine'

Summary and Conclusion

5.4. Possible Objections against Davidson's Test

Behaviorism

Mechanism

Scope of the Test

Blockheads

The Chinese Room (Internal Operation)

French: Associative Priming and Rating Games

Ability to cope with Abstract Terms

- It remains unclear how well Davidson's theory might be able to cope with abstract terms in the language.
- He throws very little light on how understanding of abstract terms might work.
- There are good reason's to doubt that communication about abstract terms can work in the same way as other communication. Something like the Wirkungsgeschichte becomes way more plausible, if we look at how culturally loaded language is
- See also Bertram's critique: Gadamer's idea

(Bertram 2012) (Davidson 1997)

Wirkungsgeschichte

6. Conclusion

6.1. Prospects of Davidson's Pragmatic Theory

While Davidson is skeptical about the possibility of explaining our linguistic performance. He believes that there is a pragmatic and empirical theory about the essence of linguistic competence.

[9] His idea that truth conditions (T-Sentences) are all that is needed to explain interpretation, and any more fundamental notion of meaning is misconceived, has also led to what is known as Davidson's Program: The attempt to give an account of the interpretation of utterances purely with the means of first order logic.

6.2. Implications for Artificial Intelligence and Computer Linguistics

Davidson tries to device a theory that can model human intellectual abilities in an empiric theory that has the power to explain them. (As opposed to only a simulation of the brain by a machine)

While it currently seems more probable that Turing's Test and also Davidson's Test will be passed by a computer that leverages trained large scale deep neural networks aka. something into the direction of simulating the brain. Davidson's consideration still provide important input into what is needed to achieve human level of natural language abilities. That is the ability to somehow interact with the world and have some sensory impression of it and have a triangular model of determining world interaction.

Possibility of AI and the importance of Language - Davidson sees artificial intelligence as possible. - He agrees that language is the key test for intelligence.

Important boundary conditions for linguistic competence - Linguistic and intellectual competence are inherently social/intersubjective - Linguistic and intellectual competence require interaction with the shared world (empiric)

Denial of conventionalism and statistical linguistics - The social and empiric traits of linguistic and intellectual competence can not be modeled by conventionalism, if this convention requires a previously shared common practice. Interpretation emerges in the instance of communication and cannot be predetermined in anyway. - Abilities of linguistic interaction can not be clearly separated from abilities for physical interaction. Both are necessary for intelligence.

References

AISB. n.d. "Loebner Prize." http://www.aisb.org.uk/events/loebner-prize.

Bertram, Georg W. 2002. "Übergangsholismus." *Zeitschrift Für Philosophische Forschung*.

Bertram, Georg W. 2012. "Antwort und Zugang." In *Hermeneutik Und Die Grenzen Der Sprache*, edited by Ulrich Arnswald, Jens Kertscher, and Louise Röska-Hardy. Heidelberg.

Chomsky, Noam. 2014. *Aspects of the Theory of Syntax*. MIT press.

Copeland, B Jack. 1999. "A lecture and two radio broadcasts on machine intelligence by Alan Turing." In *Machine Intelligence 15, Intelligent Agents [St. Catherine's College, Oxford, July 1995]*, p. 445–476. Oxford University.

Copeland, B Jack. 2000. "The Turing test." *Minds And Machines* 10 (4): p. 519–539.

Copeland, B Jack, and Diane Proudfoot. 1999. "Alan Turing's Forgotten Ideas in Computer Science." *Scientific American*, p. 99–103.

Davidson, D. 1995. "The Problem of Objectivity." *Tijdschrift Voor Filosofie* 57 (2): p. 203–220.

Davidson, Donald. 1963. "Actions, Reasons, and Causes." *Journal of Philosophy* 60 (23): p. 685–700.

Davidson, Donald. 1965. "Theories of Meaning and Learnable Languages." In *Proceedings of the International Congress for Logic, Methodology, and Philosophy of Science*, edited by Y ehoshua B ar-H illel, p. 3–17. North-Holland.

Davidson, Donald. 1967. "Truth and Meaning." *Synthese* 17 (3): p. 304–323.

Davidson, Donald. 1970. "Mental Events." In *Experience and Theory*, edited by L F oster and J W S wanson, p. 79–101. Humanities Press.

Davidson, Donald. 1973a. "On the Very Idea of a Conceptual Scheme." *Proceedings and Addresses of the American Philosophical Association* 47 (n/a): p. 5–20.

Davidson, Donald. 1973b. "Radical Interpretation." *Dialectica* 27 (1): p. 314–328.

Davidson, Donald. 1974a. "Belief and the Basis of Meaning." *Synthese* 27 (July-August): p. 309–323.

Davidson, Donald. 1974b. "Replies to David Lewis and W.V. Quine." *Synthese* 27 (3-4): p. 345–349.

Davidson, Donald. 1976. "Reply to Foster." In *Truth and Meaning: Essays in Semantics*, edited by G areth E vans and J ohn H enry M c D owell, p. 33–41. Clarendon Press.

Davidson, Donald. 1978. "What Metaphors Mean." *Critical Inquiry*, no. 5: p. 31–47.

Davidson, Donald. 1980. "Toward a Unified Theory of Meaning and Action." *Grazer Philosophische Studien* 11: p. 1–12.

Davidson, Donald. 1984. "Communication and Convention." *Synthese* 59 (1): p. 3–17.

Davidson, Donald. 1985. "Incoherence and Irrationality." *Dialectica* 39 (4): p. 345–354.

Davidson, Donald. 1986. "A Nice Derangement of Epitaphs." In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, edited by E rnest L epore, p. 433–446. Blackwell.

Davidson, Donald. 1988. "The Myth of the Subjective." In *Bewusstsein, Sprache Und Die Kuns*, edited by M K rausz, p. 221–240. Notre Dame University Press.

Davidson, Donald. 1991a. "Epistemology Externalized." *Dialectica* 45 (23): p. 191–202.

Davidson, Donald. 1991b. "Three Varieties of Knowledge." In *Royal Institute of Philosophy Supplement*, edited by A P hillips G riffiths, p. 153–166. New York: Cambridge University Press.

Davidson, Donald. 1992. "The Second Person." *Midwest Studies in Philosophy* 17 (1): p. 255–267.

Davidson, Donald. 1993. "Method and Metaphysics." *Deucalion*, no. 11: p. 239–248.

Davidson, Donald. 1994. "The Social Aspect of Language." In *The Philosophy of Michael Dummett*, edited by Gianluigi Mc Guiness Brian and Oliveri, p. 1–16. Kluwer.

Davidson, Donald. 1995. "Could There Be a Science of Rationality?" *International Journal of Philosophical Studies* 3 (1): p. 1–16.

Davidson, Donald. 1997. "Gadamer and Plato's Philebus." In *The Philosophy of Hans-Georg Gadamer*, edited by L Hahn. Chicago.

Davidson, Donald. 1999. "Spinoza's Causal Theory of the Affects." In, edited by Y irmiahu Y ovel. Little Room Press.

Davidson, Donald. 2001a. *Essays on Actions and Events: Philosophical Essays Volume 1*. Oxford University Press.

Davidson, Donald. 2001b. *Inquiries Into Truth And Interpretation: Philosophical Essays Volume 2*. Oxford University Press.

Davidson, Donald. 2001c. *Subjective, Intersubjective, Objective: Philosophical Essays Volume 3*. Clarendon Press.

Davidson, Donald. 2001d. "What Thought Requires." In *The Foundations of Cognitive Science*, edited by J oao B ranquinho, p. 121. Oxford: Clarendon Press.

Davidson, Donald. 2004. *Problems of Rationality: Philosophical Essays Volume 4*. Oxford: Oxford University Press.

Davidson, Donald. (1990a) 2004. "Representation and Interpretation." In *Modelling the Mind*, edited by K V Newton-Smith W H and Wilkes, p. 13–26. Oxford University Press.

Davidson, Donald. (1990b) 2004. "Turing's Test." In *Problems of Rationality: Philosophical Essays Volume 4*. Oxford University Press, Originally published in *Modelling the Mind*, edited by K Said, Oxford University Press: 1990.

Davidson, Donald. 2005. *Truth, Language and History: Philosophical Essays Volume 5*. Oxford University Press.

Davidson, Donald, and Gilbert Harman. 1970. "Semantics of Natural Language." *Synthese* 22 (1-2): p. 1–2.

Davidson, Donald, and Ernest Lepore. 2004. "An Interview with Donald Davidson." In *Problems of Rationality: Philosophical Essays Volume 4*. Oxford: Oxford University Press.

Descartes, René. (1637) 1993. Project Gutenberg.

Gómez-Torrente, Mario. 2015. "Alfred Tarski." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta. http://plato.stanford.edu/archives/spr2015/entries/tarski/.

Jurafsky, Daniel, and James H Martin. 2015. *Speech and Language Processing*. 2nd ed. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.

Malpas, Jeff. 2015. "Donald Davidson." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta. http://plato.stanford.edu/archives/fall2015/entries/davidson/.

Millar, P Hartley. 1973. "On the Point of the Imitation Game." *Mind* 82 (328). Mind Assoc: p. 595–597.

Moor, James. 1976. "An Analysis of the Turing Test." *Philosophical Studies* 30 (4): p. 249.

Muehlhauser, Luke. 2013. "What is AGI?" https://intelligence.org/2013/08/11/what-is-agi/.

Oppy, Graham, and David Dowe. 2011. "The Turing Test." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta. http://plato.stanford.edu/archives/spr2011/entries/turing-test/.

Schubert, Lenhart. 2015. "Computational Linguistics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta. http://plato.stanford.edu/archives/spr2015/entries/computational-linguistics/.

Searle, John R. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (3). Cambridge University Press: p. 417–424.

Shieber, Stuart M. 1994. "Lessons from a restricted Turing test." *ArXiv Preprint Cmp-Lg/9404002*.

Speaks, Jeff. 2014. "Theories of Meaning." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta. http://plato.stanford.edu/archives/fall2014/entries/meaning/.

Turing, Alan. 1950. "Computing Machinery and Intelligence." *Mind* 59 (October): p. 433–460.

Turing, Alan. (1948) 1992. "Intelligent Machinery." In *Collected Works of a. M. Turing: Mechanical Intelligence*, edited by D. C. Ince. Elsevier Science Publishers.

Turing, Alan. (1951) 1996. "Intelligent Machinery, a Heretical Theory." *Philosophia Mathematica* 4 (3): p. 256–260.

Turing, Alan. (1952) 1999. "Can Automatic Calculating Machines Be Said To Think?" In *A Lecture and Two Radio Broadcasts by Alan Turing*, edited by B J Copeland. In *Machine Intelligence* 15, p. 466–476, Oxford University Press.

Turing, Alan. (1951) 1999. "Can Digital Computers Think?" In *A Lecture and Two Radio Broadcasts by Alan Turing*, edited by B J Copeland. In *Machine Intelligence* 15, p. 462–465, Oxford University Press.

Wikipedia. n.d. "Helen Keller." https://en.wikipedia.org/wiki/Helen_Keller.

Wolfram, Steven. 2015. "Wolfram Language Artificial Intelligence: The Image Identification Project." http://blog.stephenwolfram.com/2015/05/wolfram-language-artificial-intelligence-the-image-identification-project.